AD-A068 359 MASSACHUSETTS INST OF TECH LEXINGTON LINCOLN LAB F/G 12/1
A SHORT CATALOG FOR INTERPRETING OPERATING CHARACTERISTIC CURVE--ETC(U)
JAN 79 N E LINDGREN F19628-78-C-0002
UNCLASSIFIED TN-1979-2 ESD-TR-79-21 NL

| OF |
AD
A068 359

END
DATE
FILMED
6 -79
DDC

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LINCOLN LABORATORY

A SHORT CATALOG FOR INTERPRETING OPERATING CHARACTERISTIC CURVES AND AN APPLICATION TO MULTIPLE THRESHOLD TESTING

N. E. LINDGREN

Group 32

TECHNICAL NOTE 1979-2

31 JANUARY 1979

LEXINGTON MASSACHUSETTS

ABSTRACT

In a classification rule for deciding between two possible classes, generally a single threshold test is used. If, however, one or both of the class probability density functions for the decision variable is multimodal or if the class variances are unequal, the situation may arise where it becomes desirable to use multiple thresholds to bracket several regions of the decision variable assigned to the two classes. An easy count of the number of inflection points in the operating characteristic curve generated from the single threshold case permits determination of the maximum number of thresholds that should be used.

CONTENTS

I. INTRODUCTION

In the discrimination problem where one wants to decide whether the scalar quantity x comes from class 1 or class 2, the optimal decision rule, using either the Bayes or Neyman-Pearson criterion, is a threshold test on the likelihood ratio: if $\Lambda(x) < T$ decide class 1, otherwise decide class 2, where $\Lambda(x) \triangleq p_2(x)/p_1(x)$ and $p_i(x)$ is the probability density function of x given class i, i = 1,2. In many problems Λ is a monotonic function of x so that the decision rule $\Lambda \gtrless T$ is equivalent to the threshold test $x \lessgtr x_o$ where x_o is the divide point on the x axis that separates the axis into class 1 and class 2 regions, i.e., $\Lambda(x_o) = T$. If, however, $\Lambda(x)$ is not a monotonic function of x, as can happen in the case of multimodal distribution densities or when one density function is sufficiently wide that it extends beyond both sides of the other class density function, then the decision rule $\Lambda(x) \gtrless T$ can give rise to several divide points on the x axis separating alternate class 1 and class 2 regions. The number of x axis divide points is the number of roots of the equation $\Lambda(x) = T$. One can either find which region the observed value of x falls in or one can check $\Lambda(x) \gtrless T$; the two approaches are identical.

The problem is that often one does not know $p_1(x)$ or $p_2(x)$ and therefore also $\Lambda(x)$. In the absence of such knowledge, the usual method of proceeding is to use the simplest decision rule, if $x < x_o$ decide class 1, otherwise decide class 2, where x_o is a single divide point on the axis. Of course this is equivalent to assuming the likelihood ratio is monotonic, which may or may not actually be the case. The next useful step is to plot the two kinds of decision error rates α vs β (normalized between 0 and 1) in the form of an operating characteristic (OC) curve, where the free running parameter is the divide point x_o. Here α is the leakage rate (deciding class 1 when class 2 is correct) and β is the false alarm rate (deciding class 2 when class 1 is correct). If the optimal decision rule would have been to use several divide points ($\Lambda(x)$ non-monotonic), then the OC curve obtained by using just one divide point will have various twists and bends that would not otherwise be there.

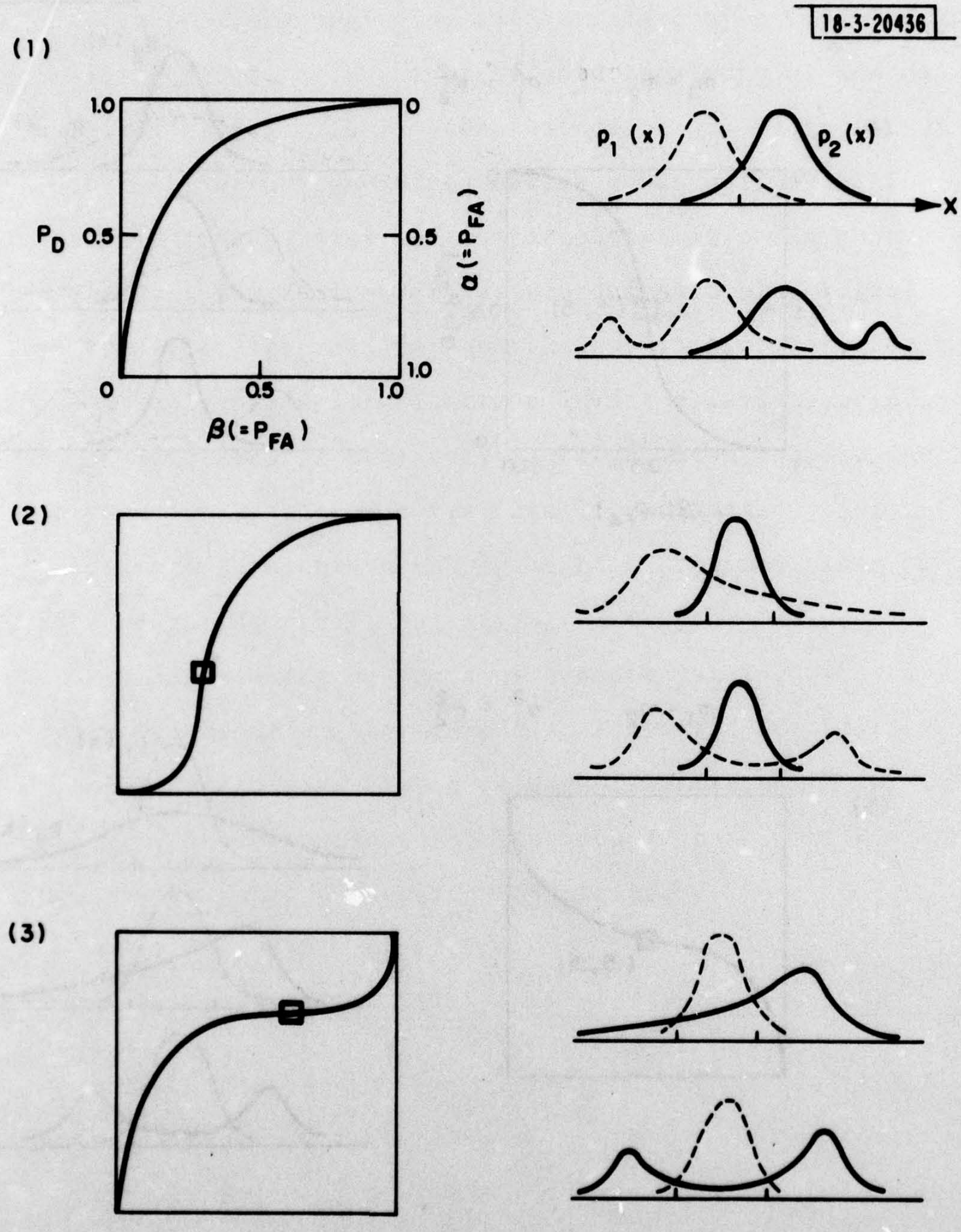
We wish to use the information contained in the twists and bends to work backwards (still in the absence of any knowledge of $\Lambda(x)$ or the class density distributions) to know how many thresholds on x, i.e., how many divide points, <u>should</u> be used in redesigning the decision rule on x if more than one threshold is appropriate. In the analysis that follows, we will get instead a weaker piece of information, namely the upper bound on the number of thresholds, based on the observed number of inflection points in the OC curve. The actual number of thresholds that should be used depends on such things as the a priori probabilities of occurrence for the two classes and the cost functions, all of which are ignored here or assumed unknown.

If one is suspicious from looking at the OC curve that $\Lambda(x)$ is non-monotonic, then one can investigate the matter more thoroughly by estimating $p_1(x)$ and $p_2(x)$ by constructing histograms or Parzen density estimates from the given data. Indeed, one could proceed from the very beginning in every case by estimating the class density functions and thereby gain knowledge of how many thresholds on x to use, but often that effort is wasted if it turns out that in fact one threshold on x was optimal. Also, estimating continuous distribution densities is not a controversy-free procedure; the choice of bin size and smoothing kernel can critically affect the results, and the question of convergence to the true underlying density function is not always assured. The suggestion advocated in this note is to go ahead and use one threshold on x, get the OC curve, and if examination of the curve shows that more than one threshold might be optimal, then construct estimates of $p_1(x)$, $p_2(x)$, and $\Lambda(x)$.

The next section contains a short catalog of example density functions and the corresponding schematic OC curves. Section III contains conclusions that have been illustrated in the catalog and proofs of those conclusions. Section IV concludes with some worked examples based on Gaussian densities.

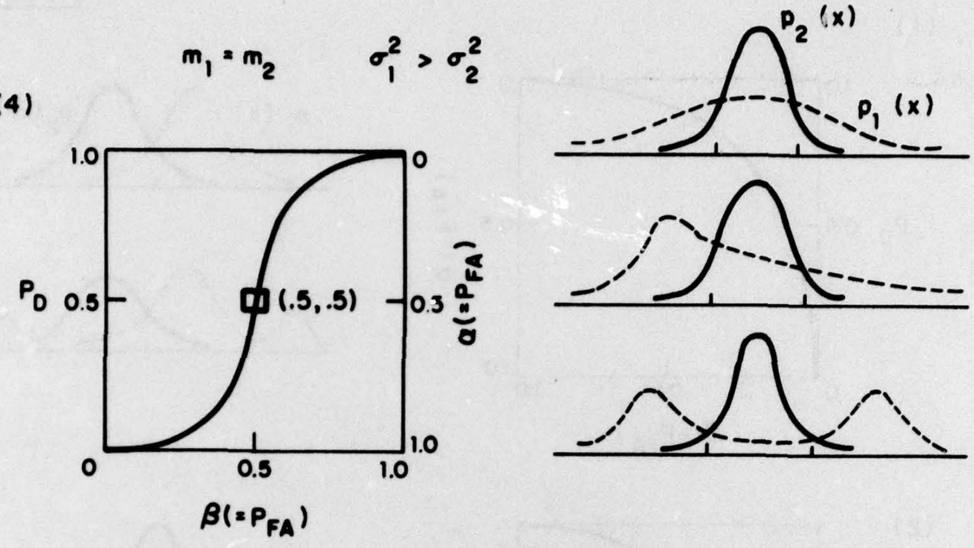
II. A SHORT CATALOG OF SINGLE THRESHOLD OC CURVES FOR VARIOUS TYPES OF CLASS DISTRIBUTIONS

The following single threshold OC curves are sketched on <u>linear</u> scales for the error rates α and β. The catalog does not begin to exhaust all the possibilities for the class distributions (probability density functions) but it should be sufficiently complete to allow the basic ideas to become apparent and to permit stating some general theorems. The OC curves are sketched based on using a single (variable) threshold x_o and the arbitrary convention chosen has been to call class 1 the class with the smaller mean value and decide class 1 if $x < x_o$, and class 2 otherwise. Scales are suppressed on all but the first OC curve. The inflection points are marked with □'s, and will be featured in a theorem later. Example (desired) decision thresholds on x are marked on the distribution sketches. Modes that occur in the wings of the distributions either do or do not affect the decision rule depending on whether they call for the addition of another threshold. Those which do not are not important and are ignored after the first catalog entry.

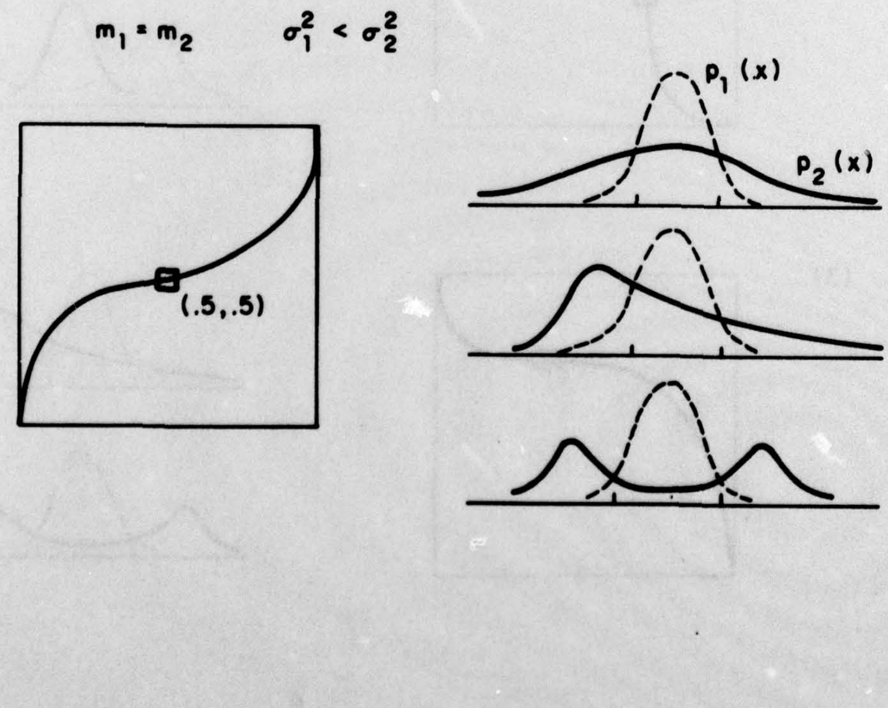

18-3-20436
(1)
1.0
0
P_D
0.5
0.5
$\alpha (= P_{FA})$
1.0
0
0.5
1.0
$\beta (= P_{FA})$
$p_1(x)$
$p_2(x)$
X
(2)
(3)

18-3-20437

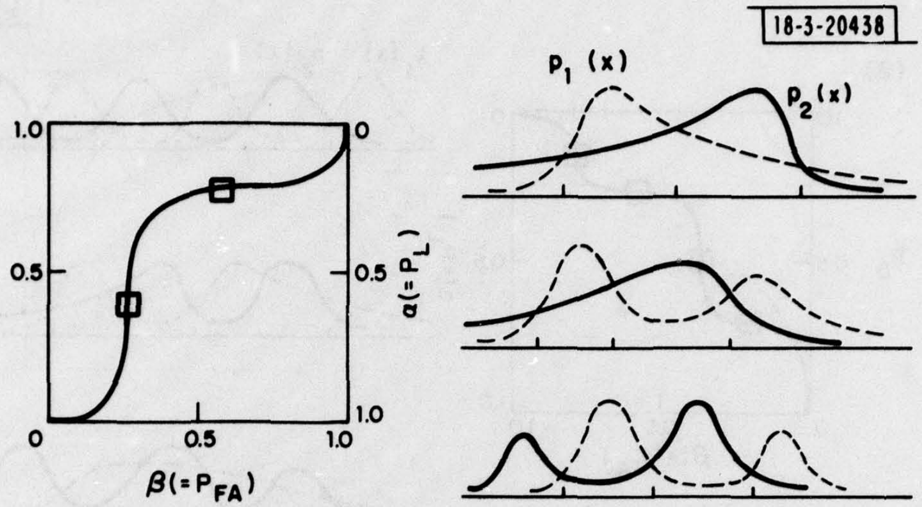
(4)


m1 = m2
σ1² > σ2²
1.0
0
PD 0.5
(.5, .5)
0.3
α(= PFA)
1.0
0
0.5
1.0
β(= PFA)
p2 (x)
p1 (x)

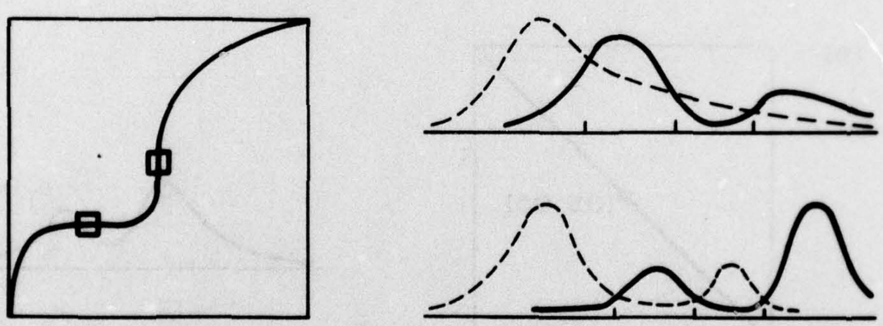
(5)


m1 = m2
σ1² < σ2²
(.5, .5)
p1 (x)
p2 (x)

(6)

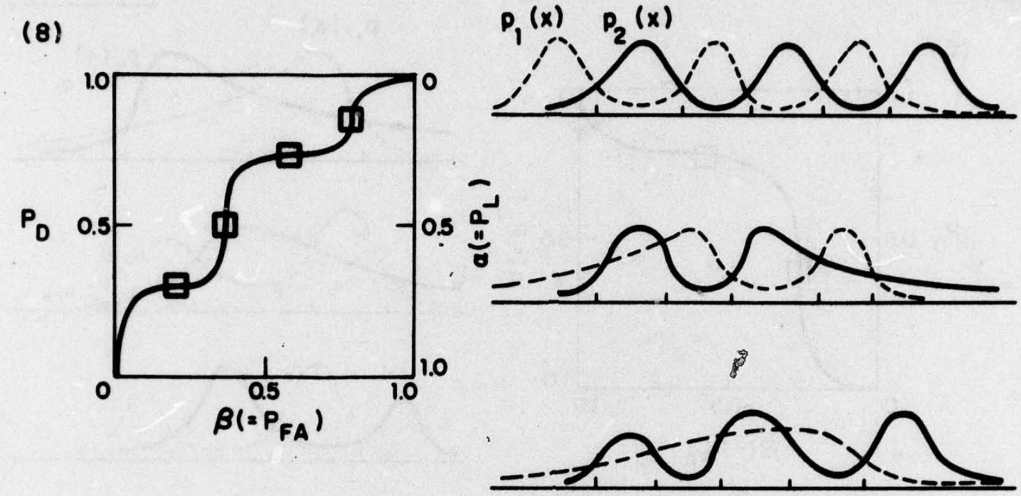


(7)

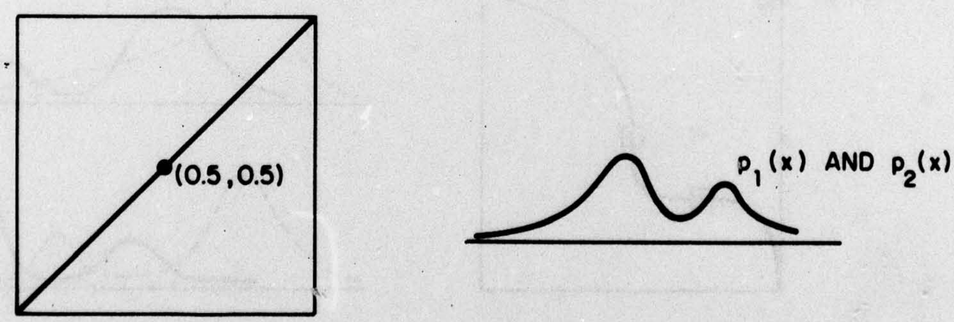


18-3-20439

(8)



(9)



III. CONCLUSIONS AND PROOFS

From the OC curve illustrations, all sketched by using a single (variable) threshold on x and linear error rate scales, the following conclusions can be made:

1. The single threshold is optimal if there is no inflection point on the OC curve.

2. More generally, the upper bound on the number of thresholds that should be used is one more than the number of inflection points on the OC curve.

3. If there is one inflection point or more on the OC curve, then the likelihood ratio, $\Lambda(x) \triangleq p_2(x)/p_1(x)$, is either (i) not a monotonic function of x, or (ii) there is a horizontal inflection point in $\Lambda(x)$, i.e., $\Lambda'(x_1) = \Lambda''(x_1) = 0$ for some x_1. (The occurrence of (ii) is expected to be very rare.)

4. The two class distributions can be considered identical if the OC curve passes sufficiently near to the equal error point (.5, .5), and the curve is sufficiently close to a 45^o angle straight line.

Proofs

Since 1. above is a special case of 2., it is sufficient to prove that 2. is true. One can begin by obtaining the condition for an inflection point in the OC curve. The detection rate is $P_D(x_o) = 1 - \alpha(x_o) = 1 - \int_{\infty}^{x_o} p_2(x)\, dx$ and the false alarm rate is $\beta(x_o) = \int_{x_o}^{\infty} p_1(x)\, dx$. The slope of the OC curve is

$$\frac{dP_D}{d\beta} = -\frac{d\alpha}{d\beta} = -\frac{d\alpha/dx_o}{d\beta/dx_o} = \frac{p_2(x_o)}{p_1(x_o)} = \Lambda(x_o), \text{ having}$$

used Leibniz's rule for differentiating the definite integrals. Differentiating again,

$$\frac{d^2P_D}{d\beta^2} = \frac{d\Lambda}{d\beta} = \frac{d\Lambda/dx_o}{d\beta/dx_o} = -\frac{1}{p_1{}^2}\left(\frac{dp_2}{dx_o} - \frac{p_2}{p_1}\,\frac{dp_1}{dx_o}\right).$$

The condition for an inflection point, $d^2P_D/d\beta^2 = 0$, can be written as $p_1'(x)/p_1 = p_2'(x)/p_2$, where the dummy argument x_o has been changed back to x, and primes denote differentiation with respect to the argument. The condition can be rewritten as

$$\frac{d}{dx}\ln\Lambda(x) = 0 \tag{1}$$

The number of inflection points in the OC curve is the number of roots of Eq. (1). Since ln Λ is a monotonic function of Λ, the number of roots and the value of the roots of Eq. (1) are identical to those of the equation

$$\Lambda'(x) = 0. \tag{2}$$

Thus the number of inflection points in the OC curve is the number of places the slope of $\Lambda(x)$ is zero. Generally that is the number of extrema of a non-monotonic $\Lambda(x)$ but the rare occurrence of an inflection point of $\Lambda(x)$ having zero slope can also contribute an OC curve inflection point. Conclusion #3 is therefore proved. By knowing the values of x at the OC curve inflection points, one then knows the values of x for which the likelihood ratio has zero slope, i.e., one knows something about $\Lambda(x)$ without having had to estimate the class density functions.

The optimal number of thresholds on x is the number of real roots of $\Lambda(x) = T$. At most, this number of roots, t, is one greater than the number of local extrema, N, in $\Lambda(x)$. The argument is topological and is illustrated for a case of N = 4 in Fig. 1 for several different values of T.

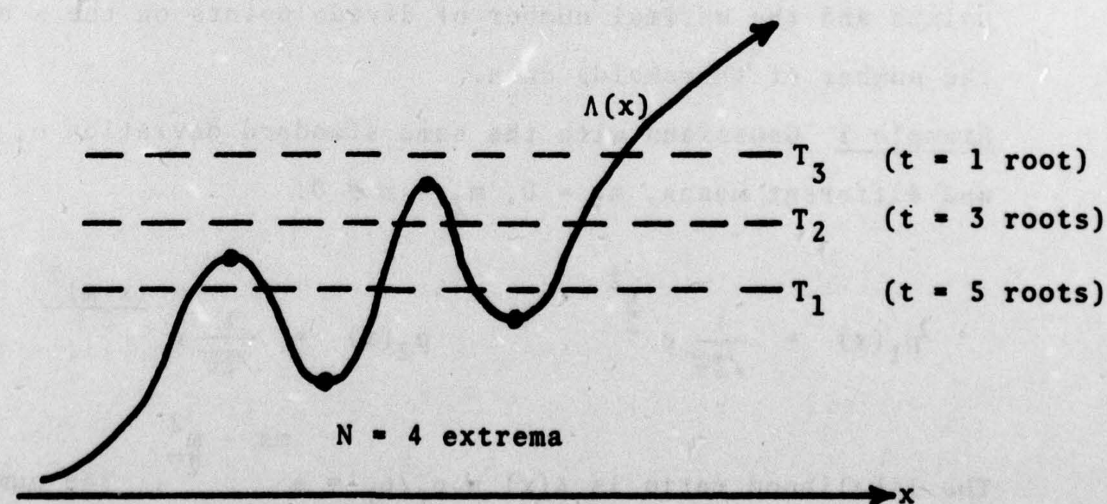


Fig. 1

In the illustration since there are 4 extrema of $\Lambda(x)$, the maximum number of roots of $\Lambda(x) = T$ is 5, as in the case of T_1. The general case is $t \leq N + 1$, and since the upper bound on N is the number of roots of $\Lambda'(x) = 0$, it follows that the upper bound on the optimal number of thresholds on x is one greater than the number of inflection points on the OC curve, proving conclusions #1 and #2.

Finally, in conclusion #4, if $p_1(x) = p_2(x)$ for all x, then $\alpha = 1 - \beta$ and the resulting OC curve is a 45^o straight line, a well known result.

IV. EXAMPLES USING GAUSSIAN DENSITY FUNCTIONS

To conclude this note, two examples are offered, illustrating the relation between the number of OC curve inflection points and the optimal number of divide points on the x axis, i.e., the number of thresholds on x.

Example 1 Gaussians with the same standard deviation $\sigma_1 = \sigma_2 = 1$ and different means, $m_1 = 0$, $m_2 = m \neq 0$:

$$p_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad p_2(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2}}$$

The likelihood ratio is $\Lambda(x) = p_2/p_1 = e^{mx - \frac{m^2}{2}}$. The number of inflection points on the OC curve is given by the number of roots of $\Lambda'(x) = 0$, or more conveniently here by Eq. (1),

$[\ln \Lambda(x)]' = 0$. In this case it is the equation $m = 0$, an equation which contradicts the initial assumption, $m \neq 0$, and therefore the equation is not satisfied for any x, i.e., it has no roots and therefore the OC curve has no inflection point. Therefore we should expect one threshold to be optimal, by conclusion #1. That such is the case is found by noting $\Lambda(x)$ is a monotonic function of x and therefore $\Lambda(x) = T$ has one root, implying that one threshold on x is, in fact, optimal.

Example 2 Gaussians with the same mean, m, but different standard deviations, $\sigma_1 = 1$, $\sigma_2 = \sigma \neq 1$.

$$p_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2}} \qquad p_2(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

The likelihood ratio is $\Lambda(x) = \frac{1}{\sigma} e^{-\frac{1}{2}\left(\frac{1}{\sigma^2} - 1\right)(x - m)^2}$.

Eq. (1) becomes $[\ln \Lambda(x)]' = -\left(\frac{1}{\sigma^2} - 1\right)(x - m) = 0$, which has one root, $x = m$, so that there is one inflection point on the OC curve. Therefore by conclusion #2 we should expect the optimal number of thresholds on x to be either 2 or 1. The optimal number of thresholds is the number of roots of $\Lambda(x) = T$ and since $\Lambda(x)$ is a Gaussian function, there are two roots for meaningful T.

ACKNOWLEDGMENT

I wish to thank Dr. Robert W. Miller for reviewing this note and making several helpful suggestions.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

19

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER 18 ESD-TR-79-21	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) 6 A Short Catalog for Interpreting Operating Characteristic Curves and an Application to Multiple Threshold Testing.	5. TYPE OF REPORT & PERIOD COVERED 9 Technical Note,
	6. PERFORMING ORG. REPORT NUMBER Technical Note 1979-2
7. AUTHOR(s) 10 Nathan E. Lindgren 14 TN-1979-2	8. CONTRACT OR GRANT NUMBER(s) 15 F19628-78-C-0002
9. PERFORMING ORGANIZATION NAME AND ADDRESS Lincoln Laboratory, M.I.T. P.O. Box 73 Lexington, MA 02173	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Project No. 8X363304D215 16
11. CONTROLLING OFFICE NAME AND ADDRESS Ballistic Missile Defense Program Office Department of the Army 5001 Eisenhower Avenue Alexandria, VA 22333 12 20p.	12. REPORT DATE 11 31 Jan[redacted]79
	13. NUMBER OF PAGES 22
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Electronic Systems Division Hanscom AFB Bedford, MA 01731	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

operating characteristics
likelihood ratio
discrimination
hypothesis testing
multiple threshold testing

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In a classification rule for deciding between two possible classes, generally a single threshold test is used. If, however, one or both of the class probability density functions for the decision variable is multimodal or if the class variances are unequal, the situation may arise where it becomes desirable to use multiple thresholds to bracket several regions of the decision variable assigned to the two classes. An easy count of the number of inflection points in the operating characteristic curve generated from the single threshold case permits determination of the maximum number of thresholds that should be used.

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

207 650